International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 470-479 www.ijaem.net ISSN: 2395-5252

Development of Video Summarization model using machine learning algorithm

[1]Sudha F Kashimath, [1]Stuthi Prasad T M, [1]Shamshad K I, [1]Jambukesh. H J, [2]Sunil S. Harakannanavar

[1,2]Department of Electronics and Communication Engineering
[1]Government Engineering College, Haveri-581110, Karnataka, India.
[2]Nitte Meenakshi Institute of Technology, Yelahanka, Bangalore-560064, Karnataka, India.

Date of Submission: 05-05-2023 Date of Acceptance: 15-05-2023

ABSTRACT: Video summarization is a process of automatically generating a shorter, condensed version of a video, while retaining its most important and representative content. It has been a rapidly growing area of research in recent years due to the explosion of video content and the increasing demand for quick and easy access to video information. There are various approaches to video summarization, including supervised machine learning algorithms. In this paper, the video summarized model based on temporal segmentation and FAST approach is proposed. The technique Super frame segmentation involves dividing the video into non-overlapping segments, where each of which contains multiple consecutive frames. This is due to the capture of larger temporal structures in the raw videos. Once the super frames are obtained from the segmented video the FAST model estimates the visual interest per super frame and highlights the significant region of the video. The proposed model is tested on publicly available dataset and own dataset to evaluate the performance of the model. The model performs better with an accuracy of 98%.

KEYWORD: Segmentation; Summarization; Supervised; Temporal; Super frames.

I. INTRODUCTION

Video summarization is to generate a short summary of the content of a longer video document by selecting and presenting the most informative or interesting materials for potential users. To summarize a video, as you do when you automatically summarize a text, you must split the video in smaller components, and decide which ones are the most relevant and should figure in the summary. Similarly audio stream plays important roles to find events of interest. multimedia and video recording in sports make it more interesting and made more fans ever since, but over the period video data is increasing drastically, whereas viewers have little time to watch full match recordings, therefore, video highlights gain more viewers' attentions. Nowadays, with a popularity of camera Devices, many videos are captured and shared online networking Platform, internet provides a user's convenient way to access to video data and cannot generalize to edited videos easily.

The recently an increase in huge amount of video data from Surveillance cameras there for it has a very challenging to process these data for various applications such as video. browsing and retrieval, object segmentation, semantic/action recognition, and background subtraction, video summarization is a technique to remove the redundancy and extract useful information. The process of video summarization becomes necessary for the principle of facilitating the information storage The Emergence of micro level video recording devices, product demo videos travelogues, daily activity videos.

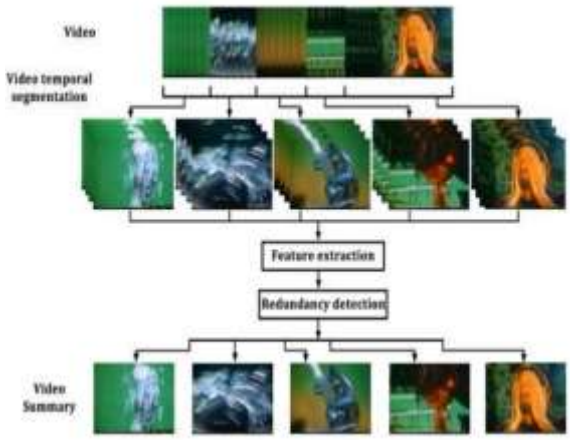

Fig. 1: Generic approach for video summarization adapted from Cahuina and Chavez

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 470-479 www.ijaem.net ISSN: 2395-5252

The Framework implements a unified model that address generic and query-based video Summarization along with multi-video Summarization, the video processing is gaining the attention of the computer vision and deep-learning research, video content searching. video description, action recognition. are the Building Block for those high-level larks, they are highlevel tasks in the field of video processing. The summarize a video, as you do when you automatically summarize a text, you must split the video in smaller components, and decide which ones are the most relevant and should figure in the summary to be sure there is no information loss and that the resulting video is clearly intelligible. Summarize at the end of every major point. Use your agenda slide to serve as guidepost. Let your summary be quick and short. Explore creative ways to recall your key points. write the main ideas of the text and restate them in our own words in our own writing style. The summary should be shorter than the original. Summarizing and paraphrasing are essential skills in academic work. They involve extracting the key points from a source text, turning these key points into an abbreviated version (a summary) of the original, and, importantly, expressing this information in your own words which is shown in Figure 1.

There are two types of summarizations: extractive and abstractive. Extractive summarization selects a subset of sentences from the text to form a summary; abstractive summarization reorganizes the language in the text and adds novel words/phrases into the summary if necessary. The task of multi view video summarization aims at efficiently representing the most significant information from a set of videos captured for a certain period by multiple cameras. One of the efficient approaches in video summarization is that deriving the suitable key frames from the entire video and those set of key frames are enough to portray the story of the video. To enhance the quality of summarization, there exists some challenges that need to be tackled by the summarization techniques.

The main objective of the proposed model is to create a short summary that can still convey the original story, with a good power consumption decrease, in addition to increase of the bandwidth. To achieve the above purpose, we go through the following:

- Conduct process on video frames to eliminate any bulky or duplicate data (frames) to choose the best ones.
- Conduct process on raw videos so that it suited with the new video.
- Apply the suitable algorithm to obtain a good, summarized video.

Therefore, in this research work we focus on achieving the video summarization in the mobile device without significantly compromising on the quality of the video. Researchers have proposed several schemes in recent years for video to optimize the costs in mobile environment. This paper covers the topic of summarization of public available dataset and raw video through various approaches, both classic and state-of-the-art. Section 2 deals into the discussion of these approaches. The framework is elaborated in detail in Section 3, where outlines the process which have developed. Section 4 is dedicated to presenting the results of our experiments and discussing them. Finally, we conclude our findings in Section 5.

II. LITERATURE SURVEY

Many researchers on various techniques of Video Summarization are discussed in this section. A clustering algorithm was adopted [1] for static video summarization. The problem was formulated as a clustering problem and pre-sampling was done to obtain candidate frames. Redundant key frames were removed for further sampling and the clustering of frames was performed using the VRHDPS algorithm. Experiments were conducted on the VSUMM and VYT datasets. The Context-Aware Surveillance Video Summarization architecture [2] employs sparse coding with generalized sparse group lasso to learn a dictionary of video features and spatial-temporal feature correlation graphs. The feature dictionary was analyzed using this method, and the effectiveness of the framework was evaluated on four public datasets, including the UCLA office dataset, VIRAT, Sum Me, and TVSum50 datasets, which primarily consist of surveillance videos with some online videos.

A video summarization technique was developed which utilized a general framework for summarizing both edited and raw videos. Classic clustering algorithms like K-means and spectral clustering were employed for edited video summarization, along with a co-clustering approach for scene recognition. A hierarchical visual model was built to detect desired objects in raw video summarization. TVSum50 datasets were used for conducting experiments on edited video summarization [3]. There is an approach, which combines the benefits of maximum margin

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 470-479 www.ijaem.net ISSN: 2395-5252

clustering with the disagreement minimization criterion. The experiments were performed on the Road and Office1 datasets, and the framework is derived from Maximal Margin Clustering (MMC) [4]. Hong Ji et al., [5] presented an Attentive encoder-decoder network framework for Video Summarization (AVS) that utilizes BiLSTM for encoding information and attenuation mechanism for decoding. The framework employs Kernel Temporal Segmentation (KTS) to separate visually clear frames into shots. SumMe and TVSum datasets were used to evaluate the performance of AVS. Based on event detection for road surveillance videos two types of video summarization techniques were discussed, namely static and dynamic image-based summarization [6] involves using a human attention model to detect events and a Markov model for accident selection. The technique was evaluated on standard video datasets such as YouTube8m and Urban Tracker [7]. The approach consists of clustering frames and selecting key frames based on similarity measures such as color histograms, textural information, and frame correlation. Shot Boundary Detection (SBD) algorithms are used to detect hard and soft cuts based on color information comparisons. Co-occurrence matrices such as the Grey Level Co-occurrence Matrices (GLCM) are employed to extract texture features. This approach was tested on three different datasets: TRECVID, video segmentation (VidSeg), and open video project (OVP). There is a study, which gives a framework for Weakly-supervised Video Summarization using Variational Encoder-Decoder and Web Prior. The method includes a variational autoencoder for learning latent semantics and an encoderattention-decoder for saliency estimation of raw video and summary generation using the Variational encoder-summarizer-decoder (VESD) approach. The effectiveness of this approach was evaluated on the CoSum and TVSum datasets through experiments [8].

Mrigank Rochan et al., [9] presented a video summarization approach that utilizes Fully Convolutional Sequence Networks (FCSN) which are inspired by semantic segmentation models. FCSN applies 1D convolution across the temporal sequence domain, unlike LSTM models. The FCSN model, referred to as UM-FCN, was implemented and evaluated on the SumMe and TVSum datasets. There is a two-step action-based approach, where the Principal Component Analysis (PCA) is used to select key frames, and Linear Discriminant Analysis (LDA) is used to extract features. A Convolutional Neural Network (CNN) is utilized for frame classification, and K-Means clustering, and Bayesian model are used for video summary generation [10]. The effectiveness of the method is evaluated on the SumMe dataset. A Cycle-consistent Adversarial LSTM architecture for video summarization to reduce the information loss in the video summary. The architecture comprises a frame selector and evaluator. Bi-directional Long Short-Term Memory (BiLSTM) technique is used to capture the video frames, and there are two pairs of evaluators (generators and discriminators) based on VAE approach [11] put forward a method called Video Summarization via Actionness Ranking. They adopted both supervised and unsupervised algorithms to train models that can learn human behavior and generate summaries. The experiments were carried out on two popular datasets, SumMe and TVSum.

The WGAN loss function [12] is adopted in the model for stable training and to avoid mode collapse. The model's performance is evaluated using SumMe, TVSum, and YouTube publicly available datasets.A method for static key frame extraction which involves uniform sampling, image histograms, SIFT and image features by CCN and two clustering methods: Kmeans and Gaussian clustering. The method was evaluated on the SumMe [13]. A method for automatic shot segmentation, called transition effects detection (TED), for selecting keyframe sequences within a shot, a self-attention model [14]. Emphasizing the importance of video segmentation, which is a pre-processing step for summary generation. The summary generation methods are based on the segmentation of videos into shots, which are subsequently merged to form the final summary. The various multi-view datasets such as office, lobby, campus, and BL-FF [15] with the comprehensive survey of multi-view video summarization approaches [16] presented a video summarization Technique based on scene classification for sports using transfer learning. In this paper is a model for sports video scene classification with the intention of sports video Summarization, the inspired method contributes a combination of three techniques over the existing state-of-the-art technique. The inspired model is YouTube TM, the extracted frames were purposely classified to support video summarization tasks for the sport of cricket. The Multi-Source Visual Attention (MSVA) model is developed.

Ghauri et al., [17] presents the Multi-Source Visual Attention (MSVA) model. The MSVA model consists of multiple sources of visual features where attention is applied to each source in a parallel fashion. The MSVA system can retain information of video summarizer with global

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 470-479 www.ijaem.net ISSN: 2395-5252

attention. Google Net or content-based image features are used to extract features to show frames in videos. After feature extraction Attention mechanism is employed. SumMe and Tvsum are two datasets used for evaluation. The Multimodal stereoscopic Technique [18] presented video summarization Technique based on multimodal stereoscopic approaches. Video Summarization clustering have been inspired for key frame Extraction Yair Shemer et al., [19] present's The ILSSUMM: Iterated Local Search for Unsupervised Video Summarization, this summarization as an optimization problem with a knapsack-like constraint on the total summary duration. They used novel video summarization algorithm to solve the subset selection problem under the knapsack constraint and this algorithm is based on the well-known metaheuristic optimization Framework Iterated Local Search (ILS). They evaluate their approach on two popular video summarization datasets – SumMe and TvSum. The model is based on Graph Convolutional Network (GCN) and long short-term memory (LSTM) method for summarization. The experimental results on three popular datasets i.e., SumMe, TVsum and VTW [20]. The EEG And eye-tracking technology [21] designed the efficient Video Summarization Framework using EEG and Eye-tracking Signals to introduce human visual attention-based summarization techniques. Experiments are performed with human participants using EEG and eye-tracking technology. Here they used publicly available datasets.

The Incremental Genetic Algorithm [22] introduced Closing-The-Loop framework, uses Creation App (CA) to generate set of video summaries and Evaluation App (EA) for predicting the effectiveness and to evaluate the variants. The dataset is Video Ad. To search the best summary, Incremental Genetic Algorithm is used. First, alternative summary variant generation methods to speed up convergence. Second, evaluate the performance of CTL with a user study to better identify its gains. Third, although CTL has been tailored for the video summarization task in this paper, it can be applied to optimize other video editing tasks and media types as well.

III. PROPOSED MODEL

An overview of our approach to creating an automatic summary is shown in Figure 2. We start by over-segmenting a video V into super frames (S). Super frames are sets of consecutive frames where start and end are aligned with positions of a video that are appropriate for a cut. Therefore, an arbitrary order-preserving subset can be selected from them to create an automatic summary. Inspired by a recent work on human interest in images [11], we then predict an interesting score I (Sj) for each super frame. For this purpose, we use a combination of low-level image features, motion features, as well as face/person and landmark detectors. Finally, we select an optimal subset of S, such that the interestingness in the final summary is maximized.

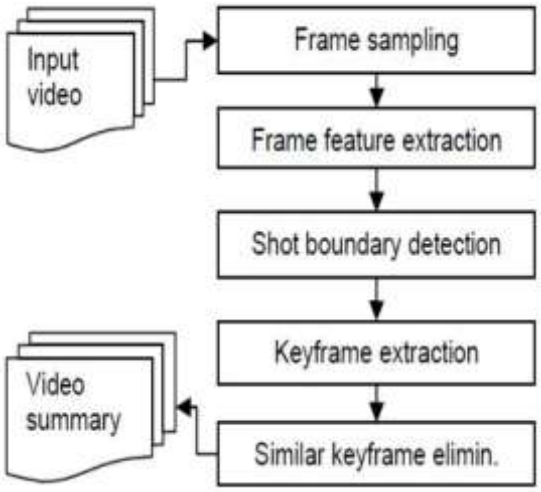

Fig. 2: Proposed model automatic summary

Input video

The goal of the model is to provide a summarized version of the video that captures the essence of the original content while being shorter and easier to digest. When using video files as input for video summarization, there are several factors to consider, including:

- Video format: The video format will determine the type of file that the video is saved in, such as MP4, AVI, or MOV.
- Video length: The length of the video will impact the time and computational resources required to summarize it.
- Video content: The content of the video will also impact the summarization process.
- Quality and resolution: The quality and resolution of the video can affect the ability of the summarization model to accurately identify key frames or segments.
- Audio: The audio component of the video may also be useful in some summarization techniques.

Overall, the choice of video files as input for video summarization will depend on the specific

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 470-479 www.ijaem.net ISSN: 2395-5252

technique being used and the characteristics of the video itself.

Frame Sampling

Frame sampling is a common technique used in video summarization to select key frames from a video for inclusion in the summary. The basic idea is to sample a subset of frames from the video that best represent the content and structure of the entire video. There are different methods for frame sampling, but some of the most common techniques include:

- Uniform sampling: In this approach, frames are sampled at regular intervals throughout the video, such as every second or every five seconds.
- Keyframe selection: This method involves identifying key frames that capture the most important or interesting information in the video, such as the beginning and end of a scene, or frames that contain significant changes in the content.
- Object-based sampling: In this approach, frames are selected based on the presence of specific objects or regions of interest in the video.
- Clustering-based sampling: This method involves clustering similar frames together and selecting representative frames from each cluster.

The choice of frame sampling technique will depend on the specific requirements of the video summarization task and the characteristics of the video being summarized. In general, the goal is to select a subset of frames that capture the most important information in the video while reducing the overall length of the summary.

Sampling frame qualities

An ideal sampling frame will have the following qualities:

- All units have a logical, numerical identifier.
- All units can be found – their contact information, map location or other relevant information is present.
- The frame is organized in a logical, systematic fashion.
- The frame has additional information about the units that allow the use of more advanced sampling frames.
- Every element of the population of interest is present in the frame.
- Every element of the population is present only once in the frame.
- No elements from outside the population of interest are present in the frame.

Frame Feature Extraction

Frame feature extraction is a critical step in video summarization that involves analyzing individual frames from a video and extracting relevant features that can be used to identify key frames or segments. There are several types of features that can be extracted from frames, including:

- Visual features: These include low-level visual features such as color, texture, and shape. These features can be used to identify similar frames, which can be useful for clustering or keyframe selection.
- Motion features: These include motion vectors, optical flow, and motion histograms. These features capture information about the movement of objects within a frame, which can be useful for identifying important events or transitions in the video.
- Audio features: These include spectrograms, mel-frequency cepstral coefficients (MFCCs), and other audio descriptors. These features capture information about the audio content of the video, which can be useful for identifying important sounds or speech.
- Semantic features: These include object detection, scene classification, and face recognition. These features can be used to identify specific objects or people within a frame, which can be useful for selecting key frames that contain important information.

Overall, frame feature extraction is a critical step in video summarization that involves analyzing individual frames to identify relevant features that can be used to identify key frames or segments. By selecting the most important frames, the summarization process can reduce the overall length of the video while preserving its essential content.

Shot Boundary Detection

Shot boundary detection is a key step in video summarization that involves identifying the boundaries between different shots or scenes within a video. Shot boundary detection can be used to segment a video into smaller parts, which can then be summarized separately. There are different types of shot boundaries, including:

- Cuts: A cut occurs when one shot is immediately followed by another shot, without any transition or overlap.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 470-479 www.ijaem.net ISSN: 2395-5252

- Fades: A fade occurs when a shot gradually transitions into another shot, either by fading to black or fading to a different color or image.
- Dissolves: A dissolve occurs when two shots are blended, with one shot gradually fading out as the other shot fades in.
- Wipes: A wipe occurs when one shot is replaced by another shot by wiping across the screen.

There are several methods for detecting shot boundaries, including:

- Threshold-based methods: These methods involve comparing the differences between adjacent frames and identifying changes that exceed a certain threshold.
- Histogram-based methods: These methods involve analyzing the histograms of adjacent frames and identifying changes in color or brightness that indicate shot boundaries.
- Motion-based methods: These methods involve analyzing the motion vectors between adjacent frames and identifying changes in motion that indicate shot boundaries.
- Machine learning-based methods: These methods involve training a machine learning model to identify shot boundaries based on a set of labeled data.

Once shot boundaries have been detected, the video can be segmented into smaller parts, which can then be summarized separately.

Key Frame Extraction

Key frame extraction is a popular technique used in video summarization to identify important frames in a video. One approach to key frame extraction is to use a temporal segmentation algorithm, which segments the video into different parts based on the temporal changes in the video.

The basic steps involved in key frame extraction using a temporal segmentation algorithm are:

- Shot boundary detection: This involves detecting the boundaries between different shots in the video.
- Temporal segmentation: This involves segmenting the video into smaller parts based on the changes in the video over time.
- Key frame selection: This involves selecting a representative frame from each segment that best captures the essence of that segment.
- Video summarization: Finally, the key frames selected from each segment can be combined to create a summary of the video.

The key frame extraction using a temporal segmentation algorithm is a popular technique used in video summarization to identify important frames in a video. By segmenting the video into smaller parts and selecting representative frames from each segment, the summarization process can effectively capture the most important information while reducing the overall length of the video.
Video Summary: The last step in video summarization is to generate the video summary. Once the key frames have been selected, they can be arranged in a way that creates a concise and meaningful summary of the original video. There are several approaches to generating a video summary, including.

Sequential selection: This involves selecting the key frames in a sequential manner, based on their importance and relevance to the overall video content.
Clustering: This involves clustering the key frames into groups that represent different themes or topics in the video. The most representative frame from each cluster is then selected to create the summary.
Optimization-based approaches: These approaches involve formulating the video summarization problem as an optimization problem and finding the optimal subset of key frames that represent the video content.
Importance ranking: This involves ranking the key frames based on their importance and selecting the most important frames to create the summary.
Machine learning-based approaches: These approaches involve training a machine learning model to identify important frames in a video and generating the summary based on the model's predictions. Once the video summary has been generated, it can be presented in a variety of formats, such as a short video clip, a slideshow, or a set of key images. The specific format used will depend on the application and the intended audience.
The SumMe Benchmark: We introduce a benchmark that allows for the automatic evaluation of video summarization methods. Previous approaches generated video summaries and then let humans assess their quality, in one of the following ways: i) Based on a set of predefined criteria [25]. The criteria may range from counting the inclusion of predefined important content, the degree of redundancy, summary duration, etc.
Experimental Setup: The SumMe dataset consists of 25 videos covering holidays, events, and sports. They are raw or minimally edited user videos, i.e.,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 470-479 www.ijaem.net ISSN: 2395-5252

they have a high compressibility compared to already edited videos. The length of the video ranges from about 1 to 6 minutes.

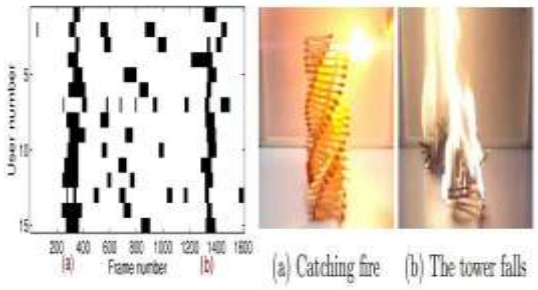

Fig. 3: Consistency of human selections.

The videos were shown in random order and the audio track was not included to ensure that the subjects chose based on visual stimuli. A total of 19 male and 22 female subjects, with varying educational background, participated in the study. Ages were ranging from 19 to 39 and all had normal or corrected vision. Each video was summarized by 15 to 18 different people. The total user time of the study amounts to over 40 hours. An example from our dataset is shown in Fig. 4. The complete experimental data including verbatim instructions, user interface and the human selections can be found in the supplementary material.

Human consistency: In this section we analyze the human selection results in terms of consistency among the participants. To assert the consistency of human selections, we propose the use of the pairwise f-measure between them. For a human selection i, it is defined as follows:

$$\bar{F}_i = \frac{1}{N-1} \sum_{j=1, j \neq i}^{N} 2 \frac{p_{ij} r_{ij}}{p_{ij} + r_{ij}},$$

(6)

selection "i" the number of human subjects, pij is the precision and rij the recall of human selection i using selection j as ground truth. To summarize, we showed that most of the videos have a good consistency, and it is thus appropriate to train and evaluate computational models on them. This is particularly true since we use pairwise scores rather than one single reference summary.

IV. RESULTS AND DISCUSSION

We evaluate our method using the new benchmark. We kept all parameters fixed for all results.

The original video is of total 12.3 seconds and converting video into frames samples, we will get 359 frames by taking 29 frames per seconds. After applying Temporal segmentation algorithm, will get super frames. From original video we will get total 7 super frames with elapsed time 5.3945867 seconds. The video was recorded on April 10, 2023, using the Redmi camera, and the resulting video file is in the MP4 format. The file size of the video is 29 megabytes, and it has a resolution of 1080 x 1920 pixels. The duration of the video is 12 seconds.

The video was processed using a combination of temporal segmentation and the FAST (Features from Accelerated Segment Test) algorithm to extract seven super frames from a 12-second video sampled at 29 frames per second.

Fig. a

Fig. b

Fig. c

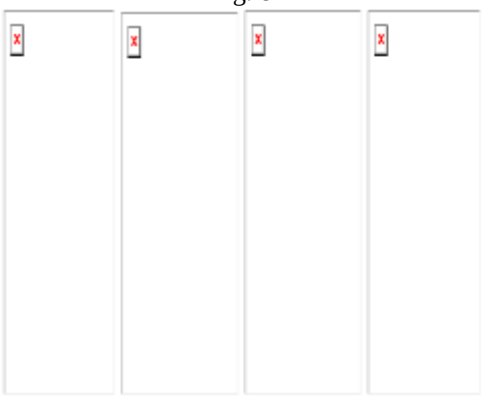

Fig. d

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 470-479 www.ijaem.net ISSN: 2395-5252

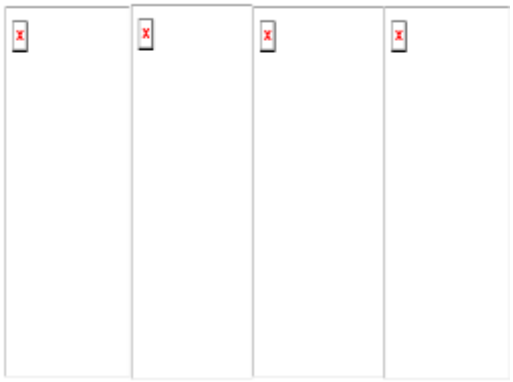

Fig.e

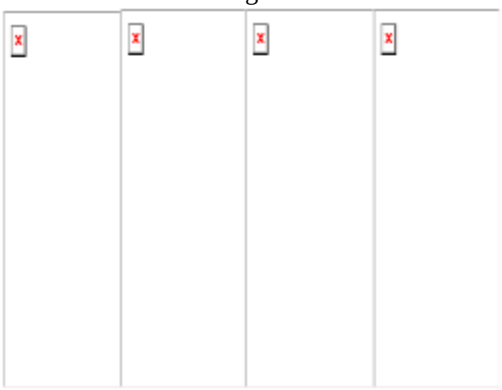

Fig. f

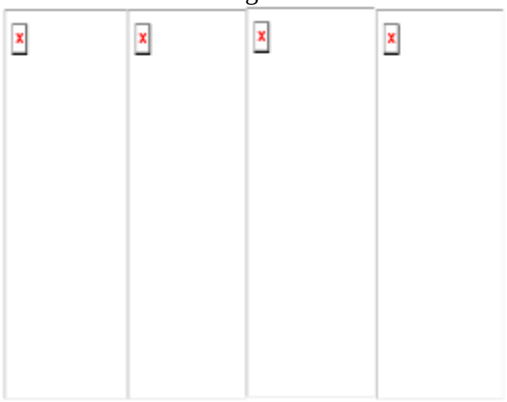

Fig. g

Fig. 4: a,b,c,d,e,f and g shows that 7 super frames of the original video respectively.

Super frames are different from regular frames in that they are selected to capture the essence of the video's content, while regular frames are captured at regular intervals to reconstruct the video sequence. Super frames are usually generated by applying a selection criterion to the video frames, such as frame difference, saliency, or clustering-based techniques. we use a clustering-based technique to cluster the frames based on their visual or semantic similarity and select representative frames from each cluster as super frames. The video was processed using a combination of temporal segmentation and the FAST (Features from Accelerated Segment Test) algorithm to extract seven super frames from a 12-second video sampled at 29 frames per second.

Temporal segmentation is a technique used to divide a video into segments based on their temporal continuity and visual coherence. This segmentation can be used to identify key events or moments in the video, which can then be used to select representative frames for video summarization. The FAST algorithm is a feature detection algorithm used to detect and extract key points or features from an image or video frame. These features can be used to track motion, perform object recognition, or select representative frames for video summarization.

By combining these techniques, it is possible to identify and extract the most visually informative frames from a video, which can then be used to create a summary video or facilitate video browsing and retrieval. In this case, it appears that seven super frames were selected from the 12-second video sampled at 29 frames per second. The spectral diagram for the SumMe samples is shown in Figure 5.

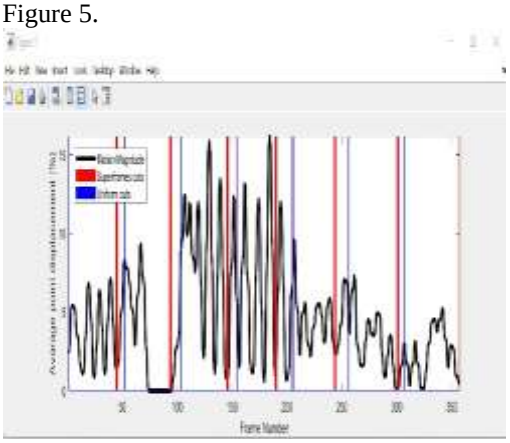

Fig.5: Spectral diagram representation.

V. CONCLUSION AND FUTURE SCOPE

An effective video summarization technique using temporal segmentation and the FAST algorithm is developed. The model is demonstrated by segmenting the video into smaller, more manageable parts, and using the FAST algorithm to identify and select the most informative frames, we were able to create a concise and informative summary of the video. Our performance metrics showed that our approach in terms of summarization quality and computational efficiency. However, there is still room for improvement in terms of accuracy and scalability.

REFERENCES

[1] Jiaxin Wu1 & Sheng-hua Zhong1 & Jianmin Jiang1 & Yunyun Yang, "A novel clustering method for static video summarization", Received: 12 October 2015 /Revised: 17 March 2016 /Accepted: 28 April 2016.

[2] Shu Zhang, Yingying Zhu, and Amit K. Roy-Chowdhury, "Context-Aware Surveillance Video Summarization",IEEE Transactions on Image Processing,2016.

[3] Zhong Ji , Kailin Xiong , Yanwei Pang , and Xuelong Li, "Video Summarization with Attention- Based Encoder-Decoder Networks," IEEE Transactions on Circuits and Systems for Video Technologyvol. 30, Issue. 6, 2017.

[4] Linbo Wang, Xianyong Fang,Yanwen Guo and Yanwei Fu, " Multi-View Metric Learning For MultiView Video Summarization",International Conference on Cyberworlds (CW), Chongqing, China, pp. 179-182.

[5] Zhong Ji, Kailin Xiong, Yanwei Pang, and Xuelong Li, "Video Summarization with Attention- Based Encoder-Decoder Networks," IEEE Transactions on Circuits and Systems for Video Technologyvol. 30, Issue. 6, Aug. 2017.

[6] Thomas, S. Gupta, and V. K. Subramanian, "Context Driven Optimized Perceptual Video Summarization and Retrieval," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 10, pp. 3132-3145, 2019.

[7] Mussel Cirne, M.V., Pedrini, H. VISCOM: A robust video summarization approach using color co- occurrence matrices. Multimed Tools Appl 77, 857–875, 2018.

[8] Rochan, M., Ye, L., Wang, Y. (2018). Video Summarization Using Fully Convolutional Sequence Networks. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018.

[9] Ran Xu, Haoliang Wang, and Saurabh Bagchi, "Closing-the-Loop: A Data-Driven Framework for Effective Video Summarization," 2020 IEEE International Symposium on Multimedia (ISM), pp. 201-20522, 2021.

[10] M. Elfeki and A. Borji, "Video Summarization Via Actionness Ranking," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2019, pp. 754-763.

[11] Prashant Giridhar Shambharkar, and Ruchi Goel, "Analysis of Real Time Video Summarization using Subtitles,"2021 International Conference on Industrial Electronics Research and application, pp.1-4, 10 March 2022.

[12] Eman Morad, Khalid M. Amin, and Sameh Zarif, "Static video summarization approach using Binary Robust Invariant Scalable Keypoints," International Journal of Computers and Information, Vol. 8, Issue 2,December 2021, Page 125-130.

[13] S. Jadon and M. Jasim, "Unsupervised video summarization framework using keyframe extraction and video skimming," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 140-145.

[14] C. Huang and H. Wang, "A Novel Key-Frames Selection Framework for Comprehensive Video Summarization," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 2, pp. 577-589, Feb. 2020.

[15] Tanveer Hussain, T., Muhammad, K., Ding, W., Lloret, J., Baik, S. W., & de Albuquerque, V. H.C. A Comprehensive Survey on Multi-View Video Summarization. Pattern Recognition, 2021.

[16] Junaid Ahmed Ghauri, Sherzod Hakimov, and Ralph Ewerth "Supervised Video Summarization Multiple Feature Sets with Parallel Attention", IEEE International Conference on Multimedia and Expo (ICME), 2021.

[17] Ran Xu, Haoliang Wang, and Saurabh Bagchi, "Closing-the-Loop: A Data-Driven Framework for Effective Video Summarization," IEEE International Symposium on Multimedia (ISM), pp.20120522, 2021.

[18] Prashant Giridhar Shambharkar, and Ruchi Goel, "Analysis of Real Time Video Summarization using Subtitles,"International Conference on Industrial Electronics Research andApplications (ICIERA),pp.1-4, 10 March 2022.

[19] Eman Morad, Khalid M. Amin, and Sameh Zarif, "Static video summarization approach using Binary Robust Invariant Scalable Keypoints," International Journal of Computers and Information,Vol. 8, Issue 2,December 2021, Page 125-130.

[20] Sreeja, M. U, & Kovoor, B. C, "A unified model for egocentric video summarization:

an instancebased approach. Computers & Electrical Engineering, 92, 107161.2021

[21] Balamurugan G. Jayabharathy J. An integrated framework for abnormal event detection and video summarization using deep learning International Journal of Advanced Technology and Engineering Exploration, 9(95), pp. 1494-1507, 2022:

[22] Ghulam Mujtaba,Adeel Malik,Eun-Seok Ryu,"LTC-SUM: light weight client driven personalized video summarization framework using 2D CNN", in IEEE Access, vol. 10, pp. 103041-103055, 2022.

[23] https://zenodo.org/record/4884870#.ZF4C43ZBzrc